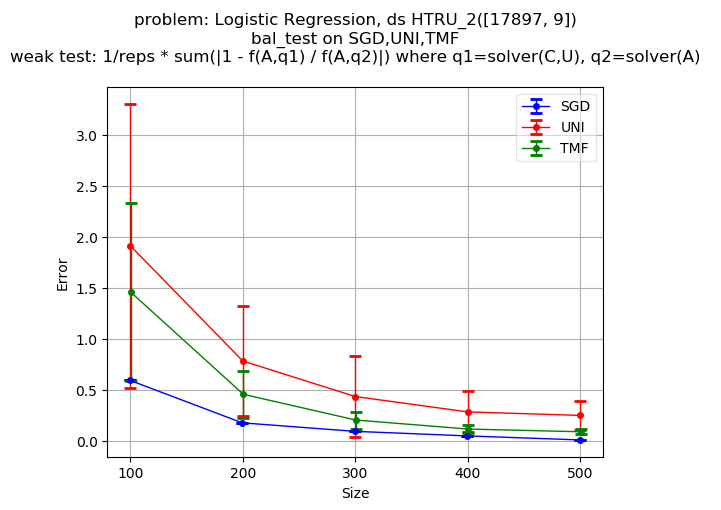
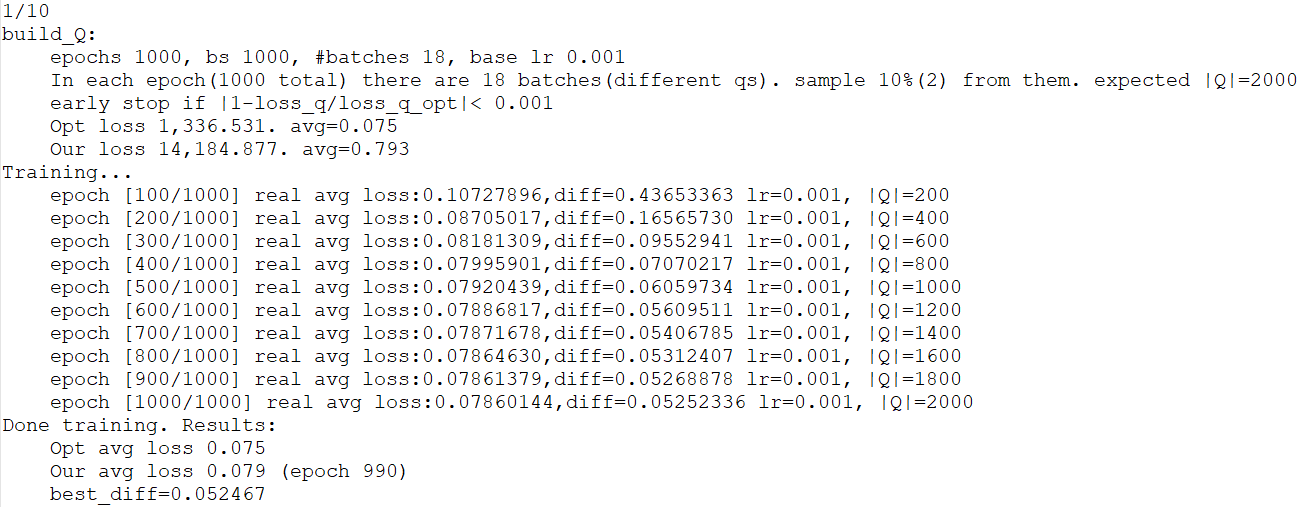
# Logistic Regression

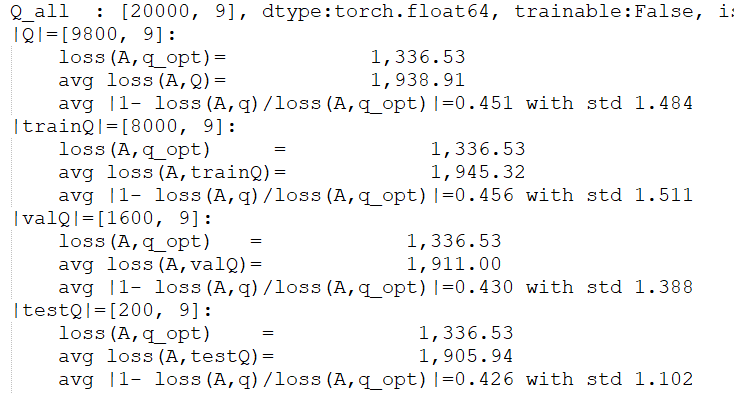
* HTRU\_2 dataset – size 17897X9
  + Data [link](https://archive.ics.uci.edu/ml/datasets/HTRU2)
* TMF – coreset made in "Coresets for Near-Convex Functions"
  + [link](https://arxiv.org/abs/2006.05482)
* Parameters:
  + Repetitions
    - 50 for UNI (Uniform sampling)
    - 1 for SGD (our work)
    - 40 for TMF (external experiment)

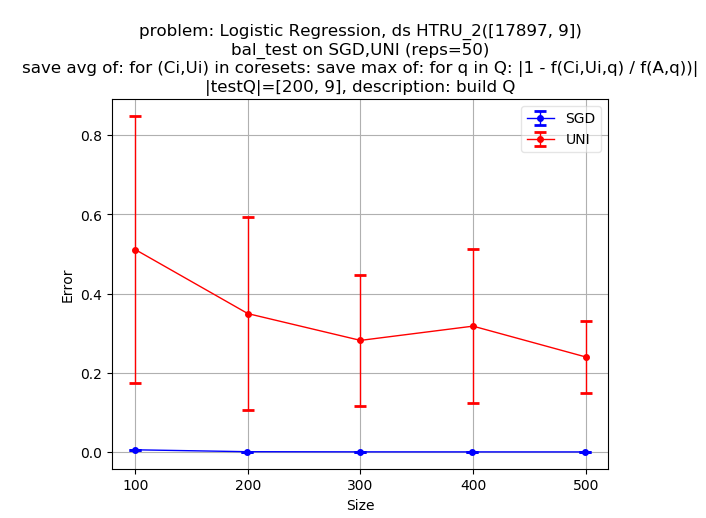


* To create a “real” Q set, we used a simple training process that finds the q opt (the optimal linear regression for this data).
  + We ran the learning process 1000 epochs
  + In each epoch we have #batches states. We sampled 10% of them.
  + We did 10 normal initializations
  + Used Adam optimizer with lr=0.001
  + Example of 1 run out of 10:



* |Q\_all| = 20000
* We sampled |Q| = 9800
  + 8000 for training
  + 1600 for validation
  + 200 for testing





## 